What Is Claimed Is:

Sub A1

~~1. A method for selecting a node to host a primary server for a service from a plurality of nodes in a distributed computing system, the method comprising:~~

~~receiving an indication that a state of the distributed computing system has changed;~~

~~in response to the indication, determining if there is already a node hosting the primary server for the service; and~~

~~if there is not already a node hosting the primary server, selecting a node to host the primary server based upon rank information for the nodes.~~

~~2. The method of claim 1, wherein selecting the node to host the primary server involves:~~

~~assuming that a given node from the plurality of nodes in the distributed computing system hosts the primary server,~~

~~communicating rank information between the given node and other nodes in the distributed computing system, wherein each node in the distributed computing system has a unique rank with respect to the other nodes in the distributed computing system,~~

~~comparing a rank of the given node with a rank of the other nodes in the distributed computing system, and~~

~~if one of the other nodes in the distributed computing system has a higher rank than the given node, disqualifying the given node from hosting the primary server.~~

3. The method of claim 2, further comprising, if there exists a node that is configured to host the primary server, allowing the node that is configured to host the primary server to communicate with other nodes in the distributed computing system in order to disqualify the other nodes from hosting the primary server.

4. The method of claim 2, wherein assuming that the given node hosts the primary server involves:

maintaining a candidate variable in the given node identifying a candidate node to host the primary server; and

initially setting the candidate variable to identify the given node.

5. The method of claim 1, further comprising, after a new node has been selected to host the primary server, if the new node is different from a previous node that hosted the primary server, establishing connections for the service to the new node.

6. The method of claim 1, further comprising, after a new node has been selected to host the primary server, if the new node is different from a previous node that hosted the primary server, configuring the new node to host the primary server for the service.

7. The method of claim 1, further comprising restarting the service if the service was interrupted as a result of the change in state of the distributed computing system.

8. The method of claim 2, wherein the given node in the distributed computing system acts a one of:

a host for the primary server for the service;

a host for a secondary server for the service, wherein the secondary server periodically receives checkpointing information from the primary server; and

a spare for the primary server, wherein the spare does not receive checkpointing information from the primary server.

9. The method of claim 8, further comprising, upon initial startup of the service, selecting a highest ranking spare to host the primary server for the service.

10. The method of claim 8, further comprising allowing the primary server to configure spares in the distributed computing system to host secondary servers for the service.

11. The method of claim 8, wherein comparing the rank of the given node with the rank of the other nodes in the distributed computing system involves considering a host for the primary server to have a higher rank than a host for a space, and considering a host for a secondary server to have a higher rank than a spare.

12. The method of claim 2, wherein disqualifying the given node from hosting the primary server involves ceasing to communicate rank information between the given node and the other nodes in the distributed computing system.

13. A computer-readable storage medium storing instructions that when executed by a computer cause the computer to perform a method for selecting a node to host a primary server for a service from a plurality of nodes in a distributed computing system, the method comprising:

receiving an indication that a state of the distributed computing system has changed;

in response to the indication, determining if there is already a node hosting the primary server for the service; and

if there is not already a node hosting the primary server, selecting a node to host the primary server based upon rank information for the nodes.

14. The computer-readable storage medium of claim 13, wherein selecting the node to host the primary server involves:

assuming that a given node from the plurality of nodes in the distributed computing system hosts the primary server,

communicating rank information between the given node and other nodes in the distributed computing system, wherein each node in the distributed computing system has a unique rank with respect to the other nodes in the distributed computing system,

comparing a rank of the given node with a rank of the other nodes in the distributed computing system, and

if one of the other nodes in the distributed computing system has a higher rank than the given node, disqualifying the given node from hosting the primary server.

15. The computer-readable storage medium of claim 14, wherein if there exists a node that is configured to host the primary server, the method

further comprises allowing the node that is configured to host the primary server to communicate with other nodes in the distributed computing system in order to disqualify the other nodes from hosting the primary server.

16. The computer-readable storage medium of claim 14, wherein assuming that the given node hosts the primary server involves:

maintaining a candidate variable in the given node identifying a candidate node to host the primary server; and

initially setting the candidate variable to identify the given node.

17. The computer-readable storage medium of claim 13, wherein after a new node has been selected to host the primary server, if the new node is different from a previous node that hosted the primary server, the method further comprises establishing connections for the service to the new node.

18. The computer-readable storage medium of claim 13, wherein after a new node has been selected to host the primary server, if the new node is different from a previous node that hosted the primary server, the method further comprises configuring the new node to host the primary server for the service.

19. The computer-readable storage medium of claim 13, wherein the method further comprises restarting the service if the service was interrupted as a result of the change in state of the distributed computing system.

20. The computer-readable storage medium of claim 14, wherein the given node in the distributed computing system acts a one of:

a host for the primary server for the service;

a host for a secondary server for the service, wherein the secondary server periodically receives checkpointing information from the primary server; and

a spare for the primary server, wherein the spare does not receive checkpointing information from the primary server.

21. The computer-readable storage medium of claim 20, wherein upon initial startup of the service, the method further comprises selecting a highest ranking spare to host the primary server for the service.

22. The computer-readable storage medium of claim 20, wherein the method further comprises allowing the primary server to configure spares in the distributed computing system to host secondary servers for the service.

23. The computer-readable storage medium of claim 20, wherein comparing the rank of the given node with the rank of the other nodes in the distributed computing system involves considering a host for the primary server to have a higher rank than a host for a space, and considering a host for a secondary server to have a higher rank than a spare.

24. The computer-readable storage medium of claim 14, wherein disqualifying the given node from hosting the primary server involves ceasing to communicate rank information between the given node and the other nodes in the distributed computing system.

25. An apparatus that selects a node to host a primary server for a service from a plurality of nodes in a distributed computing system, the apparatus comprising:

a receiving mechanism that is configured to receive an indication that a state of the distributed computing system has changed;

a determination mechanism that is configured to determine if there is already a node hosting the primary server for the service in response to the indication;

a selecting mechanism, wherein if there is not already a node hosting the primary server, the selecting mechanism is configured to select a node to host the primary server based upon rank information for the nodes.

26. The apparatus of claim 25, wherein, in selecting a node to host the primary server based upon rank information, the selecting mechanism is configured to:

communicate rank information between the given node and other nodes in the distributed computing system, wherein each node in the distributed computing system has a unique rank with respect to the other nodes in the distributed computing system, and to

compare a rank of the given node with a rank of the other nodes in the distributed computing system.

27. The apparatus of claim 26, further comprising a disqualification mechanism that is configured to disqualify the given node from hosting the primary server if one of the other nodes in the distributed computing system has a higher rank than the given node.

28. The apparatus of claim 26, further comprising a mechanism on the primary server that is configured to communicate with other nodes in the

distributed computing system in order to disqualify the other nodes from hosting the primary server.

29. The apparatus of claim 26, wherein the selecting mechanism is configured to:

maintain a candidate variable in the given node identifying a candidate node to host the primary server; and to

initially set the candidate variable to identify the given node.

30. The apparatus of claim 25, further comprising a connection mechanism that is configured to establish connections for the service to a new node after the new node has been selected to host the primary server, and if the new node is different from a previous node that hosted the primary server.

31. The apparatus of claim 25, further comprising a mechanism that configures a new node to host the primary server for the service, after the new node has been selected to host the primary server, and if the new node is different from a previous node that hosted the primary server.

32. The apparatus of claim 25, further comprising a restarting mechanism that is configured to restart the service if the service was interrupted as a result of the change in state of the distributed computing system.

33. The apparatus of claim 26, wherein the given node in the distributed computing system acts a one of:

a host for the primary server for the service;

a host for a secondary server for the service, wherein the secondary server periodically receives checkpointing information from the primary server; and

a spare for the primary server, wherein the spare does not receive checkpointing information from the primary server.

34. The apparatus of claim 33, further comprising an initialization mechanism wherein during initialization of the service, the initialization mechanism is configured to select a highest ranking spare to host the primary server for the service.

35. The apparatus of claim 33, further comprising a promotion mechanism on the primary server that that is configured to promote spares in the distributed computing system to host secondary servers for the service.

36. The apparatus of claim 33, wherein while comparing the rank of the given node with the rank of the other nodes in the distributed computing system, the selecting mechanism is configured to consider a host for the primary server to have a higher rank than a host for a secondary server, and to consider a host for a secondary server to have a higher rank than a spare.

37. The apparatus of claim 26, wherein the selecting mechanism is configured to cease to communicate rank information between the given node and the other nodes in the distributed computing system after the given node is disqualified by the disqualification mechanism.

38. A method for selecting a node to host a primary server for a service from a plurality of nodes in a distributed computer system, comprising:

communicating disqualification information between the node and remaining nodes in the plurality of nodes;

disqualifying the node from hosting the primary server based upon the disqualification information received from the remaining nodes.

39. The method of claim 38, wherein the disqualification information comprises a node rank information.

40. The method of claim 39, wherein the node rank for a given node is calculated using an assumption that the given node hosts the primary server.

41. The method of claim 40, wherein the calculated node rank is unique with respect to the ranks of other nodes in the distributed computer system.

42. The method of claim 39, wherein the disqualifying of the node comprises:

comparing a rank of the node to a set of ranks of the remaining nodes in the distributed computer system; and

disqualifying the node from hosting the primary server if one of the set of ranks of the remaining nodes is higher than the rank of the node.

43. The method of claim 38, further comprising repeating the acts of communicating disqualification information and disqualifying the node for at least one more node in the plurality of nodes.